Wiley
Journal of Advanced Transportation
Volume 2024, Article ID 7450495, 17 pages
https://doi.org/10.1155/atr/7450495

WILEY

Research Article

Evaluating Transit-Oriented Development Performance: An Integrated Approach Using Multisource Big Data and Interpretable Machine Learning

Huadong Chen,[1] **Kai Zhao**,[2] **Zhan Zhang**,[3] **Haodong Zhang**,[1] **and Linjun Lu**[1]

[1]*Department of Transportation Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*
[2]*HYP Architecture Co., Ltd., Shanghai 200240, China*
[3]*Department of Design, School of Design, Shanghai Jiao Tong University, Shanghai 200240, China*

Correspondence should be addressed to Zhan Zhang; zhanzhang@sjtu.edu.cn

Received 22 November 2023; Revised 6 April 2024; Accepted 26 September 2024

Academic Editor: Inhi Kim

Transit-oriented development (TOD) strategies on subway stations have been implemented in many high-density cities globally to enhance public transportation system efficiency and promote public transportation mobility. Focusing on the developments of intricate metropolitan systems, researchers attempted to elicit "latent rules" by proposing a generic TOD performance evaluation system. This study suggests a multi-indicator TOD performance evaluation method based on a multi-indicator approach grounded in the analysis of multisource urban big data, revealing the role of rail transit TOD station characteristics on critical indicators of station operation through an interpretable machine learning approach. Using Shanghai, China, as a case study, the methodology employed 26 widely used indicators related to TOD development and utilized a BP neural network model trained in a sample space of 77 rail transit TOD stations, aiming to predict the four critical station performance indicators. The robustness of the explanatory variables in the model has been verified by various methods, affirming their consistencies with the development characteristics of the city and the stations. The performance assessment methodology achieves significant predictive results and is computationally feasible, with potential values in applications in other high-density cities worldwide.

1. Introduction

1.1. Transit-Oriented Development (TOD) in a Developing Country. China has undergone rapid urbanization in recent decades, primarily unprecedented in magnitude and pace. The accelerated urbanization process has brought significant transportation challenges, such as increased congestion and environmental pollution. In response, China has embarked on constructing extensive rail transit networks. As of January 2023, the nation boasts urban rail transit systems in 54 cities, with 291 operational lines spanning 10 857.17 km, constituting 26.23% of the global urban rail transit system mileage [1]. Despite these developments, issues of congestion and related problems remain persistent. Thus, numerous Chinese cities have adopted the TOD strategy to address these challenges.

TOD is an urban development paradigm prioritizing public transportation as the cornerstone of urban planning. TOD aims to enhance urban areas' spatial configuration, strengthen public transport systems' efficiency, mitigate traffic congestion, and promote urban sustainability. This approach is characterized by high-density and mixed-use land development near railway stations.

Based on the core concept of high-density, mixed-use land utilization, and human-centric design, TOD advocates for integrated land use in areas surrounding transit stations, intending to support sustainable transportation methods, such as walking and cycling, while diminishing the dependency on private vehicles. The resultant effects are multifaceted, encompassing economic growth in adjacent regions, increased urban vitality, enhanced social connectivity, and an elevated standard of living for city inhabitants.

1.2. Advancing TOD Performance: New Challenges and Machine Learning. As an encompassing framework for urban planning, TOD promotes sustainable urban development processing, reducing urban traffic burdens, conserving land resources, and enhancing public urban space use efficiency. Moreover, TOD is pivotal in advancing the construction of dense, green, and low-carbon urban environments. With increased economic and social data, including diverse geospatial and personal travel data, studies have increasingly concentrated on assessing TOD efficacy through various indicator systems. Notably, studies based on the node-location model and functional extensions have been prominent. These models integrate the characteristics of functional linkages between transportation and land use into the research framework by emphasizing aspects such as pedestrian accessibility and developmental policies. While empirical evidence has substantiated the performance of TOD via stable linear correlations with metrics such as metro patronage and population density, the current discovery of threshold effects poses new challenges to fully comprehending the mechanisms by which TOD design factors influence location performance.

The application of machine learning (ML) methodologies in transportation research has been increasingly favored due to their proficiency in managing high-complexity datasets characterized by high-dimensionality, nonlinearity, and threshold effects. Despite these advantages, the limited number of TOD case studies and the challenges inherent in gathering comprehensive TOD-related information characteristics can constrain the predictive precision of these models. A further significant problem arises from the tendency of these models to yield robust global predictions while struggling to provide detailed explanations for individual cases. This often results in opaque "black box" outcomes, which obscure the understanding of how specific site design elements affect transit system performance. Due to the interpretability of ML in recent advancements, the issue can be addressed. Attribution methodologies such as SHapley Additive exPlanations (SHAP) have emerged to enhance local interpretability. These methods elucidate the influence of input features on model predictions, facilitating model diagnostics and extracting actionable insights [2, 3]. The integration of SHAP explanations into ML models represents a significant move toward revealing the predictive mechanisms and fostering a deeper comprehension of the intricate relationship between site design features and their impact on transit research outcomes.

1.3. Research Objectives. In summary, given the growing emphasis on nonlinear effects in empirical studies addressing complex transport issues through TOD, there is a new need to evaluate the effectiveness of TOD sites in the urbanization of developing countries. In this study, we seek to introduce a novel evaluation approach that harnesses the potential of multisource big data and the analytical capabilities of interpretable ML techniques. By thoroughly considering the nonlinear influence of design factors on station performance, this approach aims to evaluate station performance and thereby provide practical insights and tools for policymakers and urban planners to refine TOD plans. This in turn will facilitate the transition to more livable and resilient urban ecosystems. To achieve this goal, a key task is to delineate the components of station performance and the significant design factors. Another aspect is to select appropriate methodologies to interpret how design factors affect station performance, and thus assess the potential performance of sites based on their design characteristics after implementation.

1.4. Scope and Limitations. This research is systematically organized into four sections. The first section includes a comprehensive literature review on the performance of TOD, describing the origins of the 26 TOD design indicators used in our evaluation framework across four critical dimensions: population dynamics, traffic configurations, property valuation, and the interplay between work and residential spaces. The second section then explains the construction and analytical interpretation of the back-propagation (BP) neural network model, incorporating SHAP for a better understanding of the model's rationale, and conducts a comparative evaluation of its predictive power against established baseline models. The third section undertakes a case study of Shanghai's extensive rail transit network, empirically validating our hypothesized influence mechanisms against the backdrop of the city's urban development nuances. This section not only demonstrates the applicability of our methodology to the Shanghai context but also highlights its potential for wider adoption in urban planning practice in developing countries, facilitated by the strategic use of open data. The concluding section synthesizes the research findings and points out some of the patterns exhibited by TOD station construction in Shanghai, thereby providing urban planners with a theoretically sound and pragmatically viable toolkit for assessing TOD station performance.

However, the study faces limitations, including the lack of a temporal comparison and insufficient consideration of transport corridors' impact on site performance. Urban development is a lengthy process, and the development stages of various stations within a city can vary significantly, affecting the comparability of results. In addition, the use of a radius to define the study area may narrowly focus on individual site performance without adequate consideration of the interconnections between sites in high-density urban environments like Shanghai. Future research should address these limitations by incorporating time-series data to evaluate long-term TOD impacts and by considering the interplay between sites to broaden the TOD concept to include transportation corridors and larger urban contexts.

2. Literature Review

In the development of modern metropolitan areas, it is necessary to adopt planning approaches from different perspectives, among which TOD has become the preferred strategy for an increasing number of cities [4–9]. TOD is a model of urban development based on the principles of intensive, mixed, and humane design. The theoretical concept of TOD can be traced back to the 1950s in the Nordic countries when it was called public transportation-oriented development (PTOD) [10]. It was not until the 1980s that the concept was more formally identified and became increasingly linked to the urban fabric and transportation infrastructure [11, 12]. Since the 1990s, as more scholars began to pay attention to and study TOD, its theoretical underpinnings have been further clarified, and success factors have been gradually revealed, such as Calthorpe's [13]) clarification of the core concept of TOD. Meanwhile, TOD has also received attention from some important international organizations, including the World Bank, and has been applied in urban planning and formed a set of guidelines [14].

Although the definition of TOD varies in the literature, it is generally agreed that it focuses on creating mixed-use, high-density, and walkable communities around transit nodes [9, 15]. As such, TOD can be described as an approach that integrates land use and transportation planning with the goal of improving accessibility for walking, bicycling, and transit, and maximizing the efficiency of the existing public transportation services. Cevero [12, 16] proposed core principles of TOD-"3D principles," i.e., density, diversity, and pedestrian-friendly design.

In recent years, the definition of TOD has also shifted toward performance indicators, with more emphasis on functional measures such as travel, behavior, and environmental performance [10, 17, 18]. Scholars have explored a considerable number of success factors of TOD to expand the research framework of TOD, e.g., Lyu et al. [19] systematically reviewed the literature on the topic, collected 94 indicators, and categorized them into three dimensions, and Jamme et al. [20] proposed the extended 6Ds framework to summarize the work on the topic. At the practical level, the TCR102 report, published in 2004 in the United States, is typical, analyzing and studying the typical cases of TOD development in the United States, providing a comprehensive analysis of the role of the TOD model in U.S. urban development and forming a five-aspect evaluation system through the establishment of indicators.

Indicator-based methods for evaluating TOD sites have also been developed over time. The earliest one is Bertolini's [21] "node-place" model, which evaluates sites by two dimensions: transportation service capacity and functional richness of the site. Bertolini [21] evaluated sites along two dimensions: transportation service capacity and functional richness of the site, and classified five ideal site types under the two-dimensional framework of the "node-place" model. Based on this foundation, subsequent researchers have further developed the classification method and typology of TOD sites, such as Huang [22], who clustered sites in Canada into 10 types by clustering directly from indicator-level data. Some researchers have also chosen to expand the dimensions and enrich the number of TOD site classifications with more dimensions, such as Lyu [19] and Zhang [23], who introduced the "development dimension" and "pedestrian-oriented dimension" in conjunction with the "3Ds" framework, respectively, for the classification of TOD sites in China. For example, Lyu and Zhang introduced the "development dimension" and "pedestrian-oriented dimension" in the 3Ds framework to categorize TOD sites in Beijing and Shanghai, the two largest cities in China.

As part of the seminal work on TOD performance assessment [15], it is proposed to evaluate the TOD-ness, or the overall level of TOD, within walking distance of a site or area, which is more commonly calculated using the hierarchical analysis method that combines the 3Ds framework with expertise and its refinements. Hierarchical analysis and its refinements [24] to obtain the scores of each dimension and justify the weighting of the factors through a stable linear relationship between the scores and the performance of the site, and provide guidelines for designers [25, 26], such as the one presented by Higgins et al. [27], directly add the standardized values of seven indicators to evaluate the TOD-ness of a station area. On the other hand, some studies have explored the TOD design indicators from the perspective of their direct effect on the station performance, and representative works have been done to predict the metro passenger flow of a station through different types of direct demand models [28, 29], as well as the station performance of real estate prices, urban vitality, and so on. These works include several forecasting methodological explorations, including multiple linear regression classes [30, 31], as well as some for nonlinear methods [32–34].

In fact, an increasing number of researchers support the existence of nonlinear modes of action, such as threshold effects between TOD design indicators and station performance, which violate the assumptions of linear regression models [35]. In some studies, ML techniques such as dynamic decision trees [36], gradient boosting regression trees [37–40], and nonlinear methods such as random forests have achieved better predictive results. Other nonlinear ML methods also have a large exploration space but face challenges in terms of sample size, stability and accuracy, and difficulty in interpretability.

3. Methodology

There are four steps in this study: data fusion, indicators selection and measurement, station performance evaluation, and explanation. Firstly, we collect multisource big data from all railway stations in Shanghai, including urban transportation infrastructure data, building information and landmark data, real estate price data, and community demographic and economic data. All of these data came from publicly available data platforms and official statistics to ensure the authority and accuracy of the data. At the same time, we collect as comprehensive as possible the indicators that appeared in the existing studies that are considered to affect the performance of TOD as independent variable indicators, which cover the planning, design, transportation, economy, society, and environment aspects and have uniqueness in terms of connotation. Meanwhile, four indicators that can be observed only after the site is actually put into operation are selected as dependent variables, and these indicators can reflect the performance of TOD from different aspects. In the third step, we built a BP neural network model based on the sample data, interpreted the model using SHAP, and compared the results with those of other statistical learning methods. Finally, based on the results of SHAP, we visualize the degree of contribution to the TOD performance of the site on each of the independent variable indicators, and try to analyze the main advantages and problems exhibited by the site in depth, and then propose targeted improvement strategies, which can provide practical guidance for the planning and TOD development of high-density cities (Figure 1).

3.1. Study Area. Shanghai was chosen as the study area for this research for two main reasons. First, Shanghai has the largest urban rail network in the world, with an average daily passenger flow of up to 10 million and an operating length of 831 km. By 2023, Shanghai Metro had 20 operating lines with 508 stations. In addition, Shanghai Metro is rapidly expanding, with nine new lines under construction and an estimated 224 km of new track [33]. This large and expanding rail transit network makes Shanghai an ideal platform for exploring TOD evaluation methodologies.

Second, Shanghai's long history of polycentric urban planning concepts makes the design and development of its metro network distinctly polycentric. As early as the 1950s, Shanghai first incorporated the concept of polycentricity into its comprehensive planning. In the 14th Five-Year Plan for Shanghai's Comprehensive Transportation Development, the Shanghai Municipal Government emphasized station-city integration, proposed the construction of a systematic and complete transportation system, and realized the integrated development of stations and their surrounding areas. As a result, communities with their own characteristics have been formed around each metro station, in line with the basic concept of TOD.

As the functions, forms, surrounding land use, and socioeconomic indicators of Shanghai metro stations are diverse and rich, this provides sufficient samples and opportunities to design and implement the evaluation method. Different variations and types of TOD can be found in Shanghai's metro stations, which helps us to understand common problems and design targeted strategies.

By evaluating the level of TOD in different metro stations in Shanghai, we can make more targeted recommendations for urban planning in Shanghai and serve as a reference for other large cities with similar urban structures and transportation systems.

3.2. Data Source. The data used in this paper are relevant data for the year 2021 in Shanghai, as shown in Table 1:

The spatial extent of TOD stations is a widely discussed topic, and in this study, a radial distance of 500 m was used as the radius of TOD stations, a criterion that has been used in studies of other East Asian metropolises [41–43], and it is recommended by China's Ministry of Housing and Urban-Rural Development [44]. After necessary pre-processing, we fused the multisource big data based on urban roads with the station and its space within 500 m as the unit. The specific steps include:

According to the spatial location of the road folding line and intersection relationship, we generated the site of the three-dimensional road network. This road network includes underground, ground, and aboveground flyovers, intersections, pedestrian crossings, entrances and exits, and other facilities.

We integrated "Point of Interest" (POI) data, which identifies specific locations that might be of use or interest like stores, parks, or transit stations, with real estate transactions. These data were then linked to building outlines and further connected to the road network by incorporating information from building entrances and exits. The spatial configuration and infrastructure specifics of 417 rail transit stations in Shanghai were thoroughly restructured by amalgamating big data from diverse sources, as illustrated in Figure 2. Utilizing high-resolution open data enabled the evaluation of the locational attributes of TOD stations, the economic and societal metrics of their surrounding areas, and their adherence to 3D principles. In addition, it facilitated the assessment of the objective of transit-friendliness via the lens of the local road network and amenities.

Subsequently, a screening process was employed on the Shanghai rail transit stations, grounded in the core tenets of TOD development. This helped to discern the stations that resonate with the TOD philosophy and categorize them as rail transit TOD stations. It is essential to acknowledge that the advent of the TOD concept in China was relatively recent, coupled with a longstanding history of rail transit station development in Shanghai. This implies that no rail transit stations are explicitly conceived and developed under the umbrella of the TOD concept. Nevertheless, major Chinese cities often adopt high-density development approaches, prioritize public transportation, and encourage the integration of station-city designs.

Consequently, many rail transit stations have been constructed and regularly revamped to mirror the essential characteristics of the TOD concept. A compendium of 71 TOD sites was identified and segregated into five distinct TOD categories, as delineated in Table 2. These TOD sites, established between 1993 and 2020, were operational for at least one year during the data accumulation phase.

3.3. Selection of Indicators. To conduct a comprehensive and practical evaluation of TOD in Shanghai, we first conducted an extensive literature review, reviewing various research papers on TOD (Bertolini, 1999; Cervero & Murakami, 2009; Dittmar & Ohland, 2012). We collected and organized over 200 potentially relevant evaluation indicators. To ensure that the data were scientific and operational, we selected 26 indicators for the independent variables used to describe TOD. This selection was based on the principles of

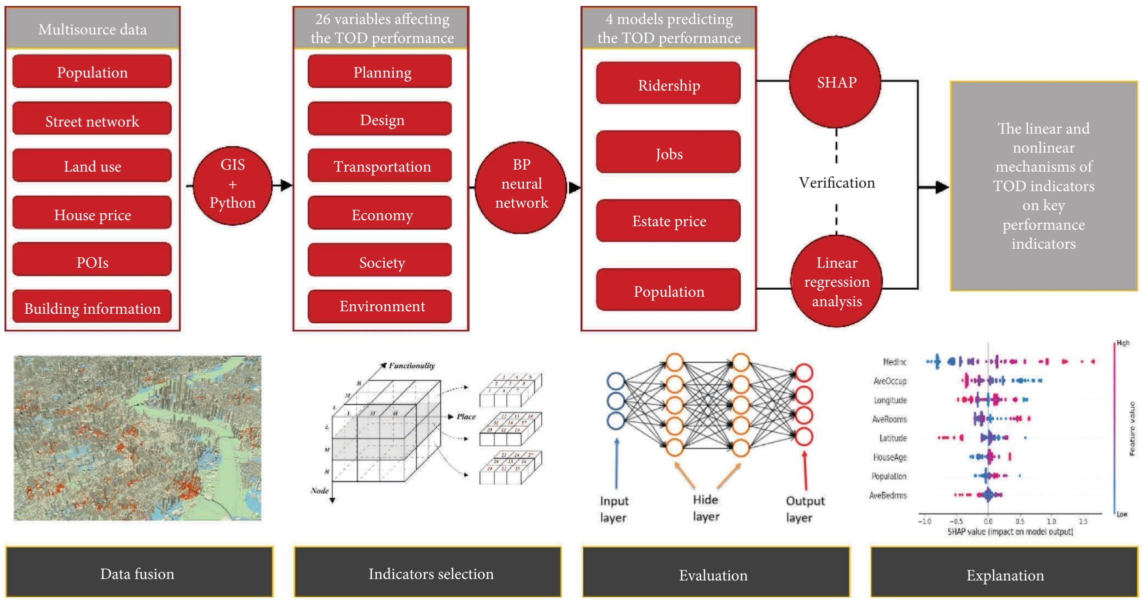


Figure 1: Overall methodological flow.

nonduplication and computability of the indicators, combined with the frequency of occurrence of the indicators in the literature. These indicators cover six dimensions: "planning," "design," "transportation," "society," "economy," and "environment", as shown in Table 3.

Meanwhile, the four dependent indicators we also selected are: metro ridership, job-housing ratio, residential prices, and population density. Compared with development density and floor area ratio, which are determined directly with the station development plan, these four indicators can be stabilized only after the station is actually put into operation, and they fit the connotative goals of TOD development, as shown in Table 4.

Figure 3 shows some of the actual data collected in this paper as an example of metro passenger flow, with the three stations with the highest passenger flow labeled. Among the TOD stations in Shanghai, the one with the highest passenger flow is Century Avenue Station, which has a daily inbound and outbound passenger flow of up to 185,653 people. This can be attributed to the fact that the station is located in the Pudong New District, one of the fastest growing and most dynamic areas in Shanghai, and that the station is a large-scale hub with four metro lines intersecting, which makes it the most centrally located node in the Shanghai metro network. The second and third ranked stations are Xinzhuang Station and Xujiahui Station, respectively, where Xinzhuang Station, as the endpoint station of Line 1 and Line 5, is an important node of the metro network connecting the Minhang District and the central city of Shanghai, while Xujiahui is one of the city centers of Shanghai.

3.4. BP Neural Network. To solve complex optimization problems in urban planning, a BP neural network model is used in this study. The BP neural network is a supervised learning algorithm used to train multilayer feed-forward neural network. The model consists of an input layer, one or more hidden layers, and an output layer. Each layer consists of multiple neurons connected by adjustable weights.

The BP neural network learning process consists of two phases: forward propagation and backpropagation. In the forward propagation phase, the input data enter the network from the input layer and is passed to the output layer after weighted summation and activation function processing layer by layer. If the output of the output layer does not match the desired output (i.e., the target value), the value of the loss function is calculated, which is usually the error between the actual output value and the target value.

In the BP stage, the error is used to calculate the gradients of each output and hidden layer neuron. These gradients indicate how the weights need to be adjusted to minimize the output error. This process is applied recursively to all neurons via the chain rule, which calculates the gradient for each weight. Updates to the weights are implemented using gradient descent or its variants, with the goal of minimizing the loss function of the entire network.

The BP neural network has the potential to become a powerful tool for urban planning problem solving. The natural nonlinear mapping ability of the BP neural network brings the possibility of capturing and learning the nonlinear relationships in urban planning data for more accurate prediction and analysis. In addition, its self-learning and self-adaptation make the BP neural network robust and fault-tolerant when the input data are noisy or partially missing, to cope with possible incomplete data situations. After proper training, the BP neural network is able to generalize and predict unseen data, which helps urban planning decision-makers to evaluate the possible impacts of different hypothetical scenarios and improves the model's ability to generalize to other cases or urban studies.

Table 1: Data source and data description.

Name	Data source	Description	Counts
Street data	Open StreetMap	Street folding data, dominated by driveways and railroads	96,911
Street data	Baidu maps API	Streets folded data of sidewalks	589,276
Building outline	Baidu maps API	Polygon shapefile for building profile,	533,207
POI (point of interest)	Gaode maps API	Point shapefile for point of interest	3,108,959
Population density	Worldpop.org	Raster dataset with image elements of 100 m size	—
Real estate price data	Lianjia.com	Records of second-hand home transactions with information about the home and its selling price	3,178,828
Metro card data	Shanghai Shentong Metro Group Co., Ltd.	Passenger arrivals and departures at Shanghai metro stations for 1 Week in August 2021	201,969

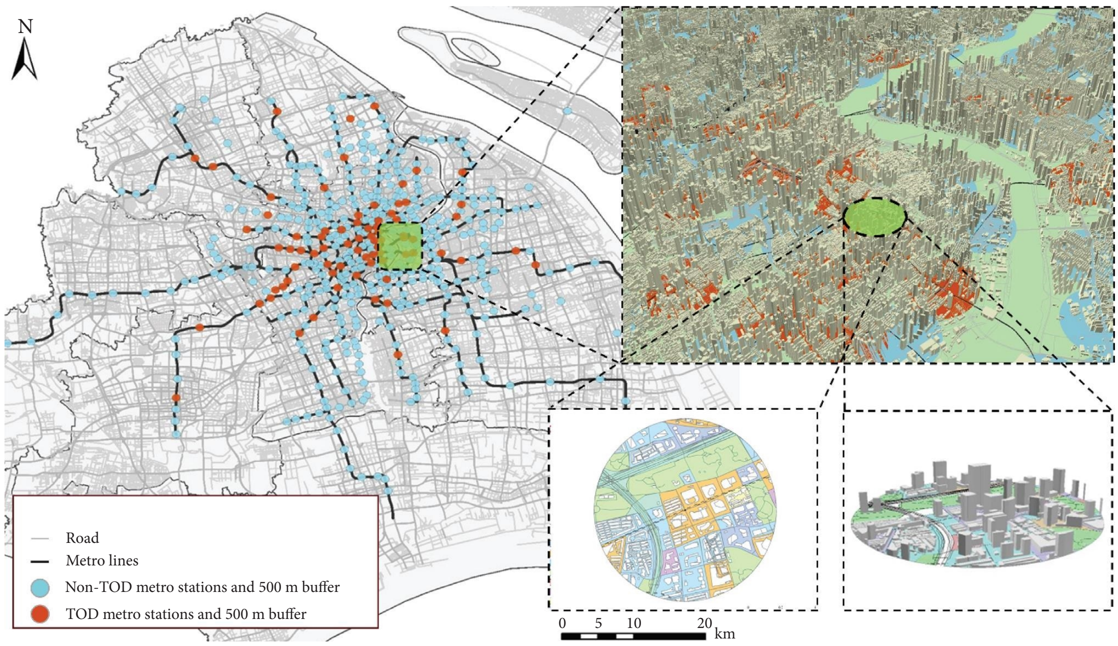


FIGURE 2: Fusion of multisource urban data and schematic distribution of transit-oriented development (TOD) stations in Shanghai.

TABLE 2: Classification results of Shanghai metro stations.

Metro line	Number of stations	Number of TOD stations	Number of non-TOD stations
1	27	8	18
2	28	11	9
3	26	4	21
4	13	4	8
5	19	1	14
6	26	3	21
7	28	4	22
8	25	10	14
9	30	6	23
10	28	5	22
11	35	7	27
12	19	1	17
13	23	2	20
14	22	2	19
15	22	0	22
16	10	0	10
17	12	2	9
18	18	1	16
Pujiang line	6	0	6
Total	417	71	346

Overall, the BP neural network is a powerful tool for urban planning problem-solving with its unique advantages. In this study, we constructed the BP neural network for each of the four performance-dependent indicators, and each neural network was set up with an input layer corresponding to the 26 features, a hidden layer of 128 neurons using the ReLU activation function, and another hidden layer of 64 neurons using the ReLU activation function. Eighty percent of our data served as the training set, and we used the mean squared error (MSE) as the loss function and the Adam optimizer to update the network weights, training each evaluation model for 1000 epochs. In addition, to verify the validity of the predictions, we also used the random forest method trained with 10-fold cross-validation and the multiple linear regression method to predict the same dataset and related metrics.

3.5. SHAP. SHAP is a game theory-based approach to explain the predictions of any ML model by decomposing the predicted values into the contributions of individual features. The SHAP values are based on the notion of equitably

Table 3: Independent variables explanation and statics.

Dimension	Indicator	Explanations	Min	Max	Mean	Std
Planning	Residential land area (km^2)	Residential land area within the station	127.5	961.6	548.9	211.0
	Commercial land area (km^2)	Commercial land area within the station	133.6	587.9	215.4	139.4
	Office land area (km^2)	Office land area within the station	0	558.9	108.1	112.1
	Administrative land area (km^2)	Administrative land area within the station	0	143.5	144.2	266.7
	School land area (m^2)	School land area within the station	0	155.5	428.3	405.3
	Hospital land area (m^2)	Hospital land area within the station	0	123.5	138.8	265.4
	Transportation land area (m^2)	Transportation land area within the station	0	674.9	546.6	867.1
Design	Average block size (m^2)	Mean value of plots of land divided by urban roads	13.8	200.6	50.3	26.1
	Community street network depth (m)	The maximum value of the shortest path between any two points in the network to describe the establishment of inter-network routes	1168.8	4733.5	1948.7	595.5
	Community street network connectivity index	The mean value of roads at each intersection in the community street network	0.882	1.395	1.070	0.076
	Community street network betweenness centrality	For node i, we define the node betweenness B_i as follows $B_i = \sum_{j \neq l \neq i} [N_{jl}(i)/N_{jl}]$, where N_{jl} is the number of shortest paths between nodes V_j and V_l, and $N_{jl}(i)$ is the number of shortest paths between nodes V_j and V_l that pass through a node V_i. $CB(V_i) = 2B_i/[(N-1)(N-2)]CB(n) = \sum_V (\max(CB(V)) - CB(V)/N - 1)$	00.093	00.518	00.283	00.096
Transportation	Number of metro station entrances and exits	The number of metro station entrances and exits within the station area	1.0	19.0	5.88	3.29
	Number of metro lines	The number of metro lines passing through stations	1.0	4.0	1.649	0.834
	Number of bus lines	The number of bus routes with bus stops set up within the station area	2.0	36.0	12.60	5.80
	Metro network betweenness centrality	The betweenness centrality indicators of rail stations in the urban rail network	0.0	0.219	0.062	0.054
	Vehicle parking capacity (vehicles)	The number of all parking spaces within the station recorded by the APP Shanghai parking	1638	8323	2238.8	1526.3
Economy	Average residential rent (yuan/m^2, monthly)	Average residential rent recorded by Lianjia.com	47.25	337.8	127.5	42.66
	Average office building rent (yuan/m^2, monthly)	Average office building rent recorded by Lianjia.com	80.0	330.0	189.3	62.40
Society	Number of large hospitals	Number of hospitals which are more than 1000 square meters within the station	0.0	6.0	0.766	1.205
	Number of large schools	Number of schools which area more than 1000 square meters within the station	0.0	11.0	2.402	2.326
	Number of large commercial facilities	Number of commercial facilities which area more than 1000 square meters within the station	0.0	9.0	2.325	1.841
	Accessibility from metro station (m)	Mean value of distances from the metro station to different types of POI by walking	127.2	981.7	456.9	133.9
	Accessibility from metro station from residential point (m)	Mean value of distances from the residential point to different types of POI by walking	0.0	981.7	79.94	144.7
	Number of business facilities	Number of facilities in the business district recorded by the Meituan APP	16	1381	323.9	325.0
Environment	Green land area (m^2)	Green land area within the station	0	481.0	539.3	910.2
	Number of large parks	Number of parks which are more than 1000 square meters within the station	0.0	9.0	0.896	1.285

Table 4: Dependent variables explanation and statics.

Indicator	Explanations	Min	Max	Mean	Std
Metro ridership	Average daily inbound and outbound passenger flow at each station	8614	185,653	53,012	34,430
Job-housing ratio	Ratio of number of jobs to residential capacity derived from POI, floor area, and related guidelines	0.836	0.243	0.487	0.201
Residential prices (yuan/m^2)	Average of selling prices per square meter for all dwellings on the station	36,232	150,952.6	86,155.2	23,630.8
Population density (person/10000 m^2)	Number of people active per 100 m size grid within the station	52.84	485.6	269.1	131.8

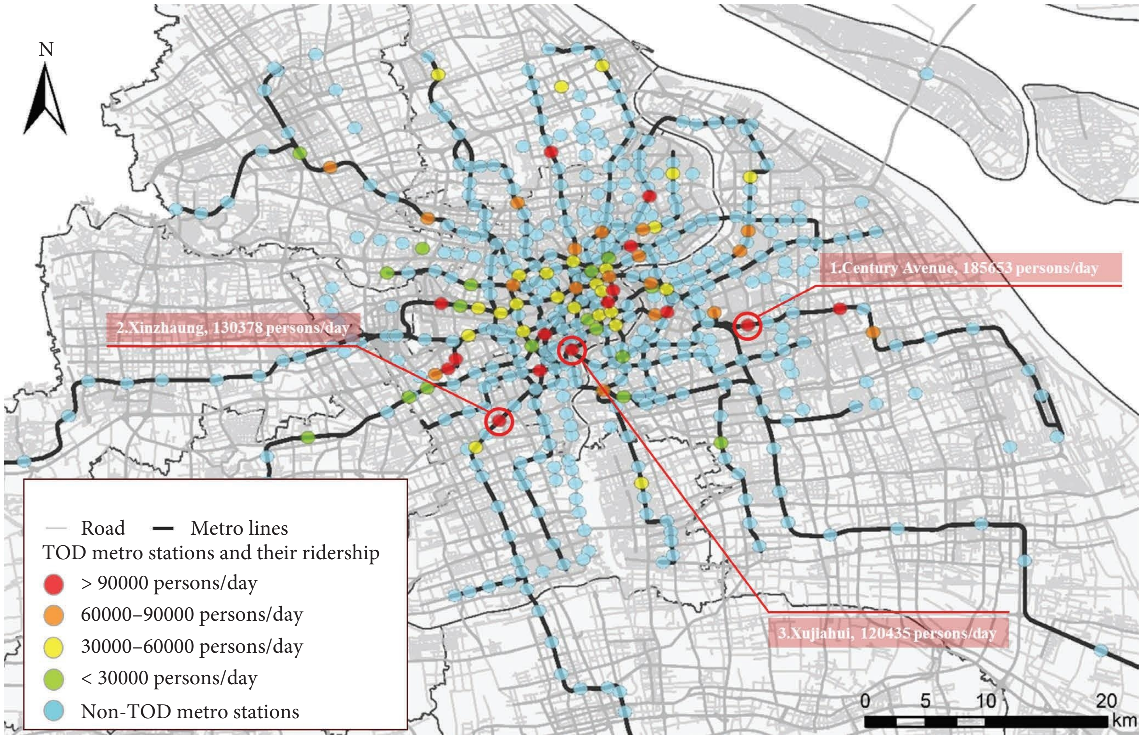


FIGURE 3: Shanghai metro passenger flow distribution.

distributing the "compensation" (predicted outputs) of the game (in this case the prediction task) among the individual "players" (features). The SHAP value is based on the notion that the "compensation" (predicted output) of the game (in this case, the prediction task) is fairly distributed to the individual "players" (features). The BP neural network, as a typical black-box model, tends to be opaque in terms of their internal decision-making process. Based on the estimation of the BP neural network, we use SHAP to explain the decision logic of the model and determine the importance of each TOD design factor on the performance index and the interaction relationship, so as to improve the interpretability and transparency of the model.

For a predictive model f, $x = (x_1, x_2, \ldots, x_n)$ is an input vector containing n features. The prediction of model f for input x can be expressed as $f(x)$. The SHAP value is intended to explain the variance of $f(x)$ with respect to the average prediction value $E[f(x)]$. For each feature x_i, its SHAP value ϕ_i is defined as:

$$\varnothing_i = \sum_{S\subseteq N\backslash\{[ii]\}} \frac{|S|!\,(n-|S|-1)!}{n!}\left(f_x(S\cup\{i\}) - f_x(S)\right), \quad (1)$$

where S is the set of features without feature i, N is the set of all features, $|S\,|$ is the size of the set S, $fx(S)$ is the predicted value of the model f when we consider only the features in the feature set S, and $|S|!\,(n-|S|-1)!/n!$ is the weight of all possible subsets of features.

4. Results and Discussion

4.1. Comparison of the BP Neural Network, Random Forest, and Regression Analysis. In this study, the relevant functions of the BP neural network, random forest, multiple linear regression, and SHAP were implemented using Python and SPSS. The prediction effects on the four indicators of metro ridership, job-housing ratio, residential prices, and population density are shown in Table 5. Overall, the BP neural network and random forest achieved a higher level of adjusted R-square than multiple linear regression, in which the BP neural network achieved higher prediction accuracy than random forest in predicting metro ridership, job-housing ratio, and population density, and the predictive ability of the two was similar in predicting residential prices. Comparing the performance indicators, the prediction accuracies of job-to-housing ratio, residential prices, and population density are higher than the metro ridership in each method, and it is worth mentioning that the prediction of residential prices is better than the rest of the performance indicators in the linear method, but this gap is significantly reduced after being processed by the nonlinear method.

Despite the low level of the adjusted R-square indicator of the multiple linear regression, a part of its detailed information can still explain the gap between the predicted effects of the model. As shown in Table 6, the independent variable indicators with higher significance and their weight values are marked, proving that there is a linear relationship between the

Table 5: Adjusted R-square of three methods for the prediction of four performance indicators.

Prediction methods	Metro ridership	Job-housing ratio	Residential prices	Population density
BP neural network	0.875	0.956	0.949	0.924
Random forest	0.700	0.869	0.920	0.849
Multiple linear regression	0.488	0.571	0.623	0.536

Table 6: Result of multiple linear regression (MLR).

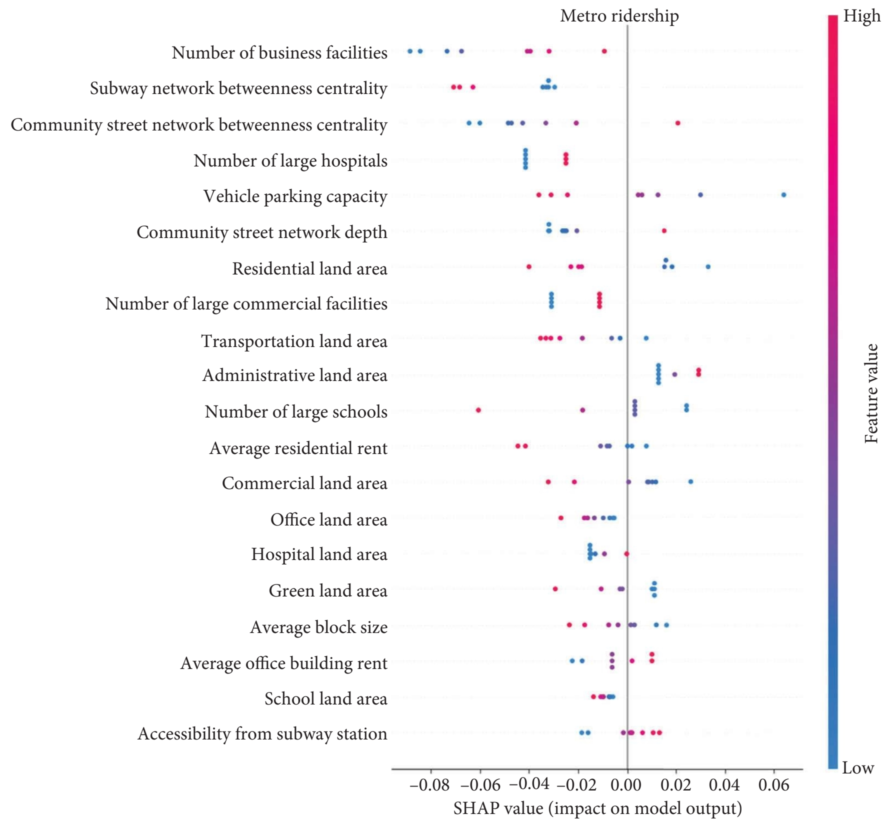
Variables	Metro ridership		Job-housing ratio		Residential prices		Population density	
	Weight	p value	Weight	p value	Weight	p value	Weight	p value
Residential land area	**−0.731**	**0.013**	−0.083	0.797	0.104	0.617	0.180	0.598
Commercial land area	**−0.696**	**0.012**	0.462	0.136	0.006	0.974	0.399	0.219
Office land area	−0.379	0.238	**0.917**	**0.015**	0.143	0.541	0.463	0.230
Administrative land area	−0.043	0.813	−0.327	0.123	−0.056	0.675	**0.391**	**0.080**
School land area	0.139	0.530	0.050	0.843	−0.046	0.776	0.067	0.801
Hospital land area	0.075	0.701	−0.026	0.907	−0.150	0.295	0.218	0.351
Transportation land area	**−0.777**	**0.050**	−0.221	0.619	−0.126	0.656	−0.221	0.636
Average block size	−0.275	0.281	−0.067	0.818	−0.027	0.886	−0.337	0.271
Community street network depth	0.131	0.480	0.154	0.470	0.114	0.401	0.079	0.721
Community street network connectivity index	−0.075	0.731	0.077	0.758	0.042	0.791	−0.061	0.814
Community street network betweenness centrality	−0.009	0.944	**−0.442**	**0.004**	−0.084	0.383	**−0.263**	**0.098**
Number of metro station entrances and exits	−0.059	0.792	−0.102	0.692	0.138	0.404	0.317	0.244
Number of metro lines	0.148	0.324	−0.140	0.413	0.053	0.625	−0.174	0.331
Number of bus lines	−0.211	0.221	0.231	0.238	0.028	0.825	**0.374**	**0.072**
Metro network betweenness centrality	0.182	0.276	0.216	0.257	−0.059	0.629	0.148	0.457
Vehicle parking capacity	−0.249	0.165	**0.355**	**0.085**	0.050	0.702	**−0.468**	**0.032**
Average residential rent	−0.143	0.655	−0.076	0.835	**0.517**	**0.031**	**0.674**	**0.083**
Average office building rent	0.118	0.527	−0.209	0.329	**0.384**	**0.007**	0.352	0.120
Number of large hospitals	0.250	0.206	−0.104	0.644	0.078	0.477	−0.055	0.814
Number of large schools	**−0.397**	**0.011**	−0.155	0.369	**0.292**	**0.046**	0.250	0.171
Number of large commercial facilities	0.164	0.351	−0.144	0.475	−0.062	0.630	−0.064	0.761
Accessibility from metro station	0.078	0.770	0.092	0.762	−0.097	0.619	0.456	0.156
Accessibility from residential point	−0.099	0.750	−0.240	0.499	−0.256	0.262	−0.297	0.425
Number of business facilities	**0.563**	**0.003**	0.015	0.945	−0.014	0.916	−0.062	0.779
Green land area	**−0.751**	**0.030**	0.487	0.211	−0.219	0.377	**0.731**	**0.076**
Number of large parks	0.142	0.648	−0.033	0.925	0.053	0.816	**−0.657**	**0.083**

Note: Bold values in the table indicate that the corresponding p value is less than 0.1, demonstrating statistical significance at the 10% level.

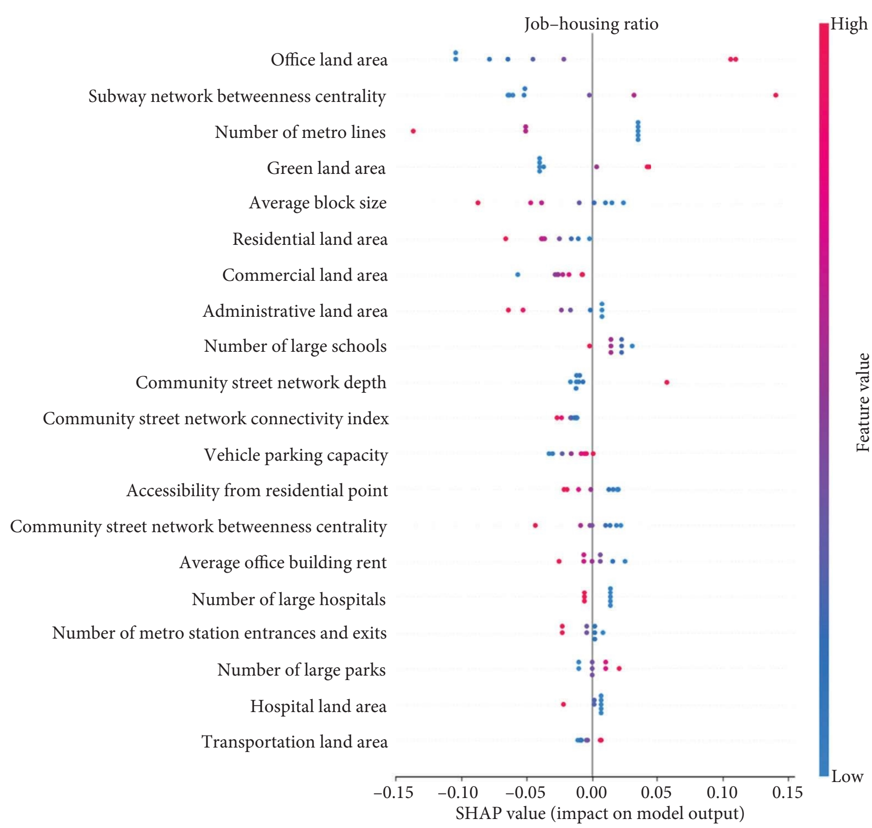
performance indicators and these independent variable indicators. The significant increase in the predictive accuracy of the jobs-housing ratio and population density in the nonlinear model may be due to the fact that the relationship between these dependent variables and the independent variables is not linear. For example, there may be threshold effects, or the relationship between the variables may be more complex (e.g., polynomial, exponential, or logarithmic), and these complex relationships cannot be captured in a multiple linear regression model. Comparatively, some important predictor variables (e.g., residential and office rent) are already included in the linear model of residential prices, and their relationships with residential prices may be inherently more linear (p values of 0.03 and 0.007, respectively), so the improvements from the nonlinear model are not as significant as for the other dependent variables.

4.2. Indicators Influences. Figure 4 presents the evaluative outcomes of the SHAP on the impact of various independent variables on the quartet of station performance metrics. The figure delineates a hierarchy of the independent variables, arranged in descending order of their contribution to the model, as discerned within each subfigure. This analysis is juxtaposed with the findings from linear regression models. The SHAP assessments corroborate the significant linear relationships identified by the linear regressions, encapsulating them within the broader interpretive framework. Furthermore, SHAP analyses illuminate the influence of additional independent variables not previously underscored, some of which exhibit pronounced nonlinear relationships. These insights reveal the nuanced complexities of the variables' contributions to station performance, enhancing the understanding of predictive dynamics within the model.

In predicting the number of metro riders, we can arrange the independent variables in the linear model in descending order of importance as follows: Number of Business Facilities (weight 0.563), Number of Large Schools (weight −0.397), Commercial Land Area (weight −0.731), Residential Land Area (weight −0.731), Green Land Area (weight −0.731), and Transportation Land Area (weight −0.731). The

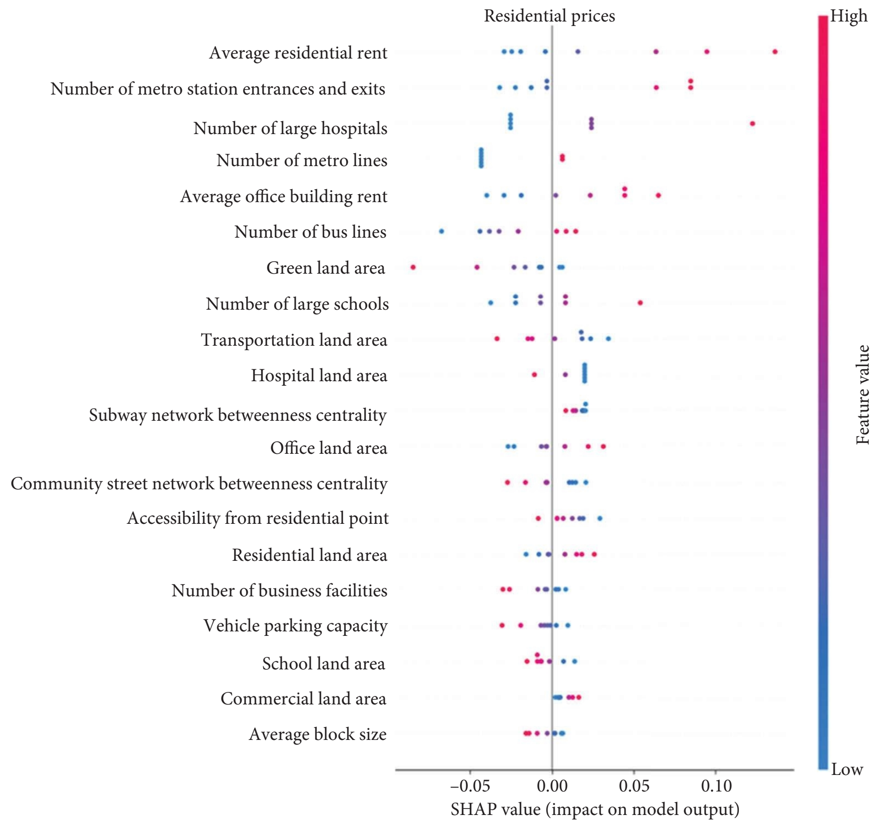


(a)

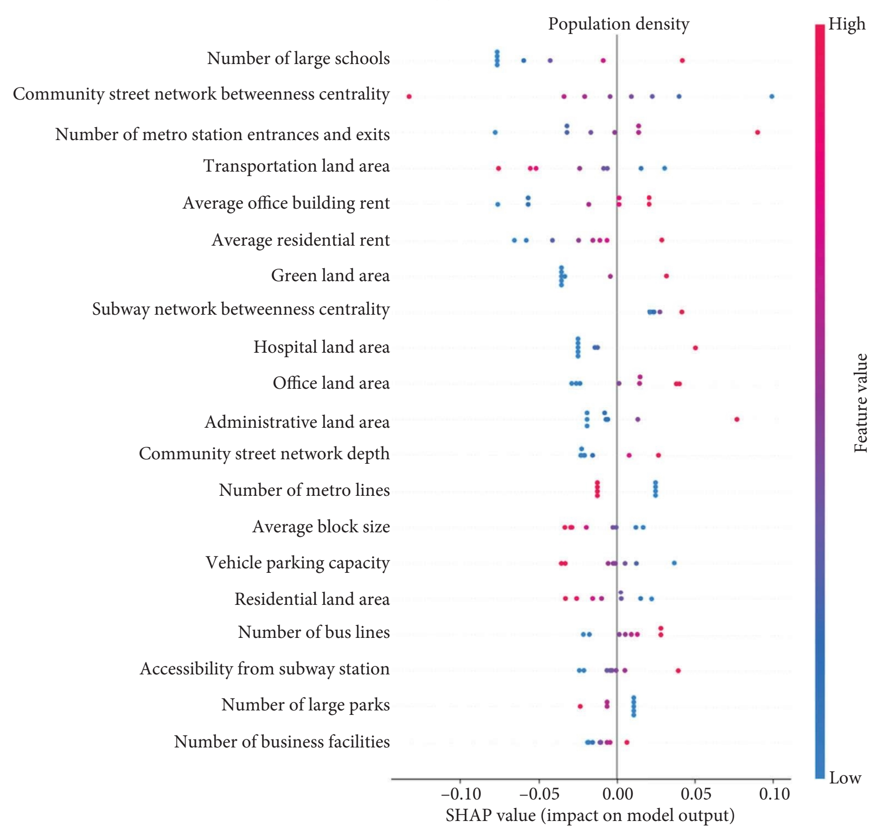


(b)

Figure 4: Continued.



(c)



(d)

Figure 4: SHAP results for different dependent variables, metro ridership (a), job-housing ratio (b), residential prices (c), and population density (d).

assigned weights to these variables are in direct correspondence with the directional changes indicated by SHAP values. Specifically, a positive weight suggests that an increase in the variable is associated with higher metro ridership. In addition, SHAP analysis identifies the positive influence of network centrality measures – Community Street Network Betweenness Centrality and Subway Network Betweenness Centrality – and the Number of Large Hospitals on ridership numbers. Conversely, the Vehicle Parking Capacity is highlighted as a negative factor. Notably, Subway Network Betweenness Centrality demonstrates a pronounced nonlinear relationship; beyond a certain threshold of the station's mediating role within the rail network, a surge in passenger flow is observed. This finding underscores the complex interplay between network centrality and ridership, warranting further investigation into the threshold effects that influence transit usage.

In predicting the job-housing ratio, the linear model finds the most significant linear relationships for Office Land Area (weight 0.917), Community Street Network Betweenness Centrality (weight −0.442), and Vehicle Parking Capacity (weight 0.355). The SHAP analysis corroborates these findings, highlighting the predominant influence of office land area, consistent with its high weight in the linear model. In addition, SHAP reveals the importance of other variables not as emphasized in the linear model, such as Subway Network Betweenness Centrality (positively correlated), Average Block Size (negatively correlated), and Residential Land Area (negatively correlated). These variables demonstrate a nonlinear impact on the job-housing ratio.

Similarly, in predicting residential prices and population density, the SHAP analysis aligns with the linear model's conclusions. Key influencing factors for residential prices include Average Residential Rent (weight 0.517), Average Office Building Rent (weight 0.384), and Number of Large Schools (weight 0.292). For population density prediction, alongside linear indicators like Green Land Area and Community Street Network Betweenness Centrality, SHAP underscores the significance of pedestrian network design elements, such as Average Block Size (negatively correlated) and Subway Network Betweenness Centrality (positively correlated). These findings highlight the multifaceted nature of factors influencing urban development metrics and underscore the utility of SHAP in revealing complex relationships within predictive models.

5. Discussion of the Evaluation Results

After the mutual verification of the linear regression model and SHAP, it is found that the predictive accuracy of a BP neural network in the context of TOD stations is significant, emphasizing the importance of understanding the relationships between independent variable indicators and performance indicators. Based on this, discussing the correlation and influence mechanism between the performance of TOD stations and TOD design indicators of Shanghai's current rail transit indicates profound values.

At the metro ridership level, the number of business facilities of rail transit stations emerges as the most influential factor in passenger flow. Continuously active commercial complexes can generate a large number of trips, attractions, and, similarly, large hospital facilities and commercial facilities. Notably, the negative impact of automobile parking capacity and land area on transit suggests that excessive transit development may diminish public transit's operational performance. In terms of community design, the positive correlation between centrally located stops on the community street network and transit usage is significant, indicating attribution to efficient connections to other buildings, enhancing public transit's appeal. In addition, the average block size indicates that minor community design can also boost transit ridership.

The study further explores the impact of indicators on the job-housing ratio, in which office land area and transportation convenience lead to more job attraction at TOD stations. Moreover, school and hospital facilities showed a high probability of skewing the site's function toward related services, while excessively high office rents are not conducive to attracting employment.

In the realm of residential prices, the SHAP methodology suggests that factors such as multiple lines, large stations with multiple entrances and exits, and proximity to large hospitals and schools align with real estate market preferences, contributing to higher residential sales prices.

Regarding population density, SHAP identifies a positive correlation with the number of large school facilities and station entrances and exits. Conversely, a higher degree of centrality within the community road network correlates negatively with population density. These findings can be contextualized within the urban fabric of Shanghai, where specific city characteristics influence residential patterns. A key factor is the local policy of schooling by neighborhood, which incentivizes parents to acquire properties in proximity to educational institutions, irrespective of the size suitability for family needs. In addition, as community density escalates, transit authorities respond by augmenting station access points to alleviate congestion during peak traffic periods. Furthermore, neighborhood stations, typically less integrated into the broader street network due to a scarcity of commercial and service centers, are characterized by a higher concentration of residential lots. These lots are often situated at a distance from central urban functions, leading to a pattern where many commuters reside in these peripheral areas while working in more central locations. This dynamic pattern contributes to the observed high population densities at these neighborhood stations.

6. Conclusions

This study aims to rigorously evaluate key performance indicators of TOD at urban rail transit stations, implementing multisource urban big data and interpretable ML methods. Focusing on Shanghai's citywide rail transit stations, which provide a unique case to examine TOD in dynamic urban environments, the research selects four critical performance indicators: metro ridership, job-housing ratio, residential prices, and population density. Employing the BP neural network and SHAP methods, the study not only predicts these indicators but also explores the mechanism of 26 TOD design indicators on station performance.

Firstly, our analysis presents a feasible approach for evaluating the performance of rail transit stations using a BP neural network. The findings indicate that this method significantly enhances prediction accuracy of the four indicators from 74.4% to 96.2% compared to traditional linear regression techniques. The BP neural network demonstrates superior capabilities in self-learning and self-adaptation, distinguishing it from other nonlinear methods; compared to the random forest approach, the method delivers a 4.97% accuracy improvement. This provides TOD researchers with a data processing methodology that uses open data sources to quickly predict and estimate station performance.

Secondly, the study validates the reliability of SHAP in elucidating the local impact within ML-based global prediction models. All 17 indicators obtained from the traditional linear regression model that have a significant impact on performance are captured by SHAP. In contrast, indicators considered less significant by the linear regression model, such as the betweenness centrality of stations in the subway network, are captured by SHAP in the form of nonlinear effects. This alignment not only substantiates the enhanced predictive capabilities of the BP neural network but also highlights the influence of nonlinear factors on site performance.

In addition, the study integrates model interpretation results with the developmental characteristics of the urban environment to investigate the mechanisms through which TOD design indicators influence rail transit station performance. For example, the SHAP model points out that the installation of large hospital facilities and school facilities within the station area is conducive for enhancing the performance of the station. The office land area of the station can only produce a significant effect of attracting jobs after exceeding a certain threshold, and higher betweenness centrality may reduce the inbound and outbound passenger flow of the station, etc. This comprehensive analysis aids in elucidating the specific mechanisms by which these design indicators impact TOD performance, offering practical insights for future TOD research and urban planning strategies.

Overall, this research introduces a data-driven approach for evaluating the performance of TOD at rail transit stations. The methodology is characterized by its interpretability and adaptability, offering an innovative perspective on the impact of station location, design, and environmental factors. This approach is particularly beneficial for urban planners, as it provides insights that can inform the timely adjustment of spatial planning, development density, and land-use strategies. The method is applicable not only in planning and designing new stations but also in updating existing ones. By facilitating these adjustments, the proposed method plays a crucial role in promoting sustainable urban development. This comprehensive approach underscores the importance of integrating data-driven analysis with practical urban planning applications, offering a pathway toward more efficient and sustainable urban growth.

Data Availability Statement

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

The authors thank to Transportation Research Center in SJTU for their technical support.

References

[1] B. Han, Z. Yang, Y. Yu, et al., "Statistical Analysis of Urban Rail Transit Operation in the World in 2020: A Review," *Urban rapid rail transit* 34, no. 1 (2021): 5–11.

[2] S. M. Lundberg, G. Erion, H. Chen, et al., "From Local Explanations to Global Understanding With Explainable AI for Trees," *Nature Machine Intelligence* 2, no. 1 (2020): 56–67, https://doi.org/10.1038/s42256-019-0138-9.

[3] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems [Advances in neural information processing systems 30 (nips 2017)]* (December 2017).

[4] S. Feldman, P. Lewis, and R. Schiff, "Transit-Oriented Development in the Montreal Metropolitan Region: Developer's Perceptions of Supply Barriers," *Canadian Journal of Urban Research* 21, no. 2 (2012): 25–44.

[5] A. Guthrie and Y. Fan, "Developers" Perspectives on Transit-Oriented Development," *Transport Policy* 51 (2016): 103–114, https://doi.org/10.1016/j.tranpol.2016.04.002.

[6] M. Kamruzzaman, F. M. Shatu, J. Hine, and G. Turrell, "Commuting Mode Choice in Transit Oriented Development: Disentangling the Effects of Competitive Neighbourhoods, Travel Attitudes, and Self-Selection," *Transport Policy* 42 (2015): 187–196, https://doi.org/10.1016/j.tranpol.2015.06.003.

[7] F. Paulhiac Scherrer, "Assessing Transit-Oriented Development Implementation in Canadian Cities: An Urban Project Approach," *Journal of Planning Education and Research* 39, no. 4 (2019): 469–481, https://doi.org/10.1177/0739456x19873942.

[8] P. Topalovic, J. Carter, M. Topalovic, and G. Krantzberg, "Light Rail Transit in Hamilton: Health, Environmental and Economic Impact Analysis," *Social Indicators Research* 108, no. 2 (2012): 329–350, https://doi.org/10.1007/s11205-012-0069-x.

[9] W.-M. Wey, H. Zhang, and Y.-J. Chang, "Alternative Transit-Oriented Development Evaluation in Sustainable Built Environment Planning," *Habitat International* 55 (2016): 109–123, https://doi.org/10.1016/j.habitatint.2016.03.003.

[10] A. Ibraeva, G. H. d. A. Correia, C. Silva, and A. P. Antunes, "Transit-Oriented Development: A Review of Research Achievements and Challenges," *Transportation Research Part A: Policy and Practice* 132 (2020): 110–130, https://doi.org/10.1016/j.tra.2019.10.018.

[11] R. Hickman and P. Hall, "Moving the City East: Explorations into Contextual Public Transport-Orientated Development," *Planning Practice and Research* 23, no. 3 (2008): 323–339, https://doi.org/10.1080/02697450802423583.

[12] R. Thomas and L. Bertolini, "Defining Critical Success Factors in TOD Implementation Using Rough Set Analysis," *Journal of Transport and Land Use* 10, no. 1 (2017): 139–154.

[13] P. Calthorpe, *The Next American Metropolis: Ecology, Community, and the American Dream* (Princeton architectural press, 1993).

[14] A. Ibraeva, B. Van Wee, G. H. d. A. Correia, and A. Pais Antunes, "Longitudinal Macro-Analysis of Car-Use Changes Resulting From a TOD-type Project: The Case of Metro Do Porto (Portugal)," *Journal of Transport Geography* 92 (2021): 103036, https://doi.org/10.1016/j.jtrangeo.2021.103036.

[15] Y. J. Singh, A. Lukman, J. Flacke, M. Zuidgeest, and M. Van Maarseveen, "Measuring TOD Around Transit Nodes-Towards TOD Policy," *Transport Policy* 56 (2017): 96–111, https://doi.org/10.1016/j.tranpol.2017.03.013.

[16] R. Cervero and K. Kockelman, "Travel Demand and the 3Ds: Density, Diversity, and Design," *Transportation Research Part D: Transport and Environment* 2, no. 3 (1997): 199–219, https://doi.org/10.1016/s1361-9209(97)00009-6.

[17] T. G. C. Shen, L. Cheng, Y. J. Yang, J. L. Deng, T. H. Jin, and M. Q. Cao, "Do Residents Living in Transit-Oriented Development Station Catchment Areas Travel More Sustainably? the Impacts of Life Events," *Journal of Advanced Transportation* 2023 (2023): 1–13, https://doi.org/10.1155/2023/9318505.

[18] Y. Singh, M. Zuidgeest, J. Flacke, and M. van Maarseveen, "A Design Framework for Measuring Transit Oriented Development," *WIT Transactions on The Built Environment* 128 (2012): 719–730.

[19] G. Lyu, L. Bertolini, and K. Pfeffer, "Developing a TOD Typology for Beijing metro Station Areas," *Journal of Transport Geography* 55 (2016): 40–50, https://doi.org/10.1016/j.jtrangeo.2016.07.002.

[20] H.-T. Jamme, J. Rodriguez, D. Bahl, and T. Banerjee, "A Twenty-Five-Year Biography of the TOD Concept: From Design to Policy, Planning, and Implementation," *Journal of Planning Education and Research* 39, no. 4 (2019): 409–428, https://doi.org/10.1177/0739456x19882073.

[21] L. Bertolini, "Nodes and Places: Complexities of Railway Station Redevelopment," *European Planning Studies* 4, no. 3 (1996): 331–345, https://doi.org/10.1080/09654319608720349.

[22] R. Huang, A. Grigolon, M. Madureira, and M. Brussel, "Measuring Transit-Oriented Development (TOD) Network Complementarity Based on TOD Node Typology," *Journal of Transport and Land Use* 11, no. 1 (2018): 305–324, https://doi.org/10.5198/jtlu.2018.1110.

[23] D. Qiang, L. Zhang, and X. Huang, "Quantitative Evaluation of TOD Performance Based on Multi-Source Data: A Case Study of Shanghai," *Frontiers in Public Health* 10 (2022): 820694, https://doi.org/10.3389/fpubh.2022.820694.

[24] C. Cucuzzella, J. Owen, S. Goubran, and T. Walker, "A TOD Index Integrating Development Potential, Economic Vibrancy, and Socio-Economic Factors for Encouraging Polycentric Cities," *Cities* 131 (2022): 103980, https://doi.org/10.1016/j.cities.2022.103980.

[25] Z. J. Cao, Y. S. Asakura, and Z. B. Tan, "Coordination Between Node, Place, and Ridership: Comparing Three Transit Operators in Tokyo," *Transportation Research Part D: Transport and Environment* 87 (2020): 102518, https://doi.org/10.1016/j.trd.2020.102518.

[26] M. A. Uddin, T. Tamanna, S. Adiba, and S. Bin Kabir, "Revolutionizing TOD Planning in a Developing Country: An Objective-Weighted Framework for Measuring Nodal TOD Index," *Journal of Advanced Transportation* 2023 (2023): 1–25, https://doi.org/10.1155/2023/9911133.

[27] C. Higgins and P. Kanaroglou, "Rapid Transit, Transit-Oriented Development, and the Contextual Sensitivity of Land Value Uplift in Toronto," *Urban Studies* 55, no. 10 (2018): 2197–2225, https://doi.org/10.1177/0042098017712680.

[28] J. Guo, F. Nakamura, Q. Li, and Y. Zhou, "Efficiency Assessment of Transit-Oriented Development by Data Envelopment Analysis: Case Study on the Den-En Toshi Line in Japan," *Journal of Advanced Transportation* 2018 (2018): 1–10, https://doi.org/10.1155/2018/6701484.

[29] R. L. Guo and Z. D. Huang, "Mass Rapid Transit Ridership Forecast Based on Direct Ridership Models: A Case Study in Wuhan, China," *Journal of Advanced Transportation* 2020 (2020): 1–19, https://doi.org/10.1155/2020/7538508.

[30] K. T. K. Nyunt and N. Wongchavalidkul, "Evaluation of Relationships between Ridership Demand and Transit-Oriented Development (TOD) Indicators Focused on Land Use Density, Diversity, and Accessibility: A Case Study of Existing Metro Stations in Bangkok," *Urban Rail Transit* 6, no. 1 (2020): 56–70, https://doi.org/10.1007/s40864-019-00122-2.

[31] A. Amini Pishro, Q. H. Yang, S. Q. Zhang, et al., "Node, Place, Ridership, and Time Model for Rail-Transit Stations: a Case Study," *Scientific Reports* 12, no. 1 (2022): 16120, https://doi.org/10.1038/s41598-022-20209-4.

[32] V. V. Ivanov and E. S. Osetrov, "Forecasting Daily Passenger Traffic Volumes in the Moscow Metro," *Physics of Particles and Nuclei Letters* 15, no. 1 (2018): 107–120, https://doi.org/10.1134/s1547477118010089.

[33] S. L. Su, Z. L. Wang, B. Z. Li, and M. J. Kang, "Deciphering the Influence of TOD on Metro Ridership: An Integrated Approach of Extended Node-Place Model and Interpretable Machine Learning With Planning Implications," *Journal of Transport Geography* 104 (2022): 103455, https://doi.org/10.1016/j.jtrangeo.2022.103455.

[34] H. T. Yang, X. Li, C. J. Li, J. H. Huo, and Y. G. Liu, "How Do Different Treatments of Catchment Area Affect the Station Level Demand Modeling of Urban Rail Transit?" *Journal of Advanced Transportation* 2021 (2021): 1–19, https://doi.org/10.1155/2021/2763304.

[35] B. van Wee and S. Handy, "Key Research Themes on Urban Space, Scale, and Sustainable Urban Mobility," *International Journal of Sustainable Transportation* 10, no. 1 (2016): 18–24, https://doi.org/10.1080/15568318.2013.820998.

[36] S. Kim, S. Rasouli, H. Timmermans, and D. J. Yang, "Estimating Panel Effects in Probabilistic Representations of Dynamic Decision Trees Using Bayesian Generalized Linear Mixture Models," *Transportation Research Part B: Methodological* 111 (2018): 168–184, https://doi.org/10.1016/j.trb.2018.03.010.

[37] C. Ding, X. Y. Cao, and C. Liu, "How Does the Station-Area Built Environment Influence Metrorail Ridership? Using Gradient Boosting Decision Trees to Identify Non-Linear Thresholds," *Journal of Transport Geography* 77 (2019): 70–78, https://doi.org/10.1016/j.jtrangeo.2019.04.011.

[38] Z. X. Gan, M. Yang, T. Feng, and H. J. P. Timmermans, "Examining the Relationship Between Built Environment and Metro Ridership at Station-To-Station Level," *Transportation Research Part D: Transport and Environment* 82 (2020): 102332, https://doi.org/10.1016/j.trd.2020.102332.

[39] Q. F. Shao, W. J. Zhang, X. Y. Cao, J. W. Yang, and J. Yin, "Threshold and Moderating Effects of Land Use on Metro Ridership in Shenzhen: Implications for TOD Planning," *Journal of Transport Geography* 89 (2020): 102878, https://doi.org/10.1016/j.jtrangeo.2020.102878.

[40] J. W. Yang, J. Cao, and Y. F. Zhou, "Elaborating Non-linear Associations and Synergies of Subway Access and Land Uses with Urban Vitality in Shenzhen," *Transportation Research Part A: Policy and Practice* 144 (2021): 74–88, https://doi.org/10.1016/j.tra.2020.11.014.

[41] J. Choi, Y. J. Lee, T. Kim, and K. Sohn, "An Analysis of Metro Ridership at the Station-To-Station Level in Seoul," *Transportation* 39, no. 3 (2012): 705–722, https://doi.org/10.1007/s11116-011-9368-3.

[42] S. L. Su, H. Zhang, M. Wang, M. Weng, and M. J. Kang, "Transit-oriented Development (TOD) Typologies Around Metro Station Areas in Urban China: A Comparative Analysis of Five Typical Megacities for Planning Implications," *Journal of Transport Geography* 90 (2021): 102939, https://doi.org/10.1016/j.jtrangeo.2020.102939.

[43] H. Sung and J. T. Oh, "Transit-oriented Development in a High-Density City: Identifying its Association with Transit Ridership in Seoul, Korea," *Cities* 28, no. 1 (2011): 70–82, https://doi.org/10.1016/j.cities.2010.09.004.

[44] Ministry of Housing and Urban-Rural Development of the People's Republic of China, "Guidelines for Planning and Design of Areas Along Urban Railways," (2015), https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/201512/20151210_225899.html.